

Posterior Consistency via Precision Operators for Bayesian Nonparametric Drift Estimation in SDEs

Y. Pokern*, A.M. Stuart†, J.H. van Zanten‡

August 28, 2012

Accepted for Publication in Stochastic Processes and their Applications

Abstract

We study a Bayesian approach to nonparametric estimation of the periodic drift function of a one-dimensional diffusion from continuous-time data. Rewriting the likelihood in terms of local time of the process, and specifying a Gaussian prior with precision operator of differential form, we show that the posterior is also Gaussian with precision operator also of differential form. The resulting expressions are explicit and lead to algorithms which are readily implementable. Using new functional limit theorems for the local time of diffusions on the circle, we bound the rate at which the posterior contracts around the true drift function.

1 Introduction

Diffusion processes are routinely used as statistical models for a large variety of phenomena including molecular dynamics, econometrics and climate dynamics (see for instance [33], [21] and [19]). Such a process can be specified via the drift and diffusion functions of a stochastic differential equation driven by a Brownian motion W . Even in one dimension, this class of processes attracts great applied interest. In this case, provided the diffusion function σ is known and under mild additional assumptions, one can transform the process such that the diffusion function is constant:

$$dX_t = b(X_t)dt + dW_t. \quad (1.1)$$

*Department of Statistical Science, University College London. Email: y.pokern@ucl.ac.uk

†Department of Mathematics, University of Warwick, Coventry CV4 7AL, England. Email: a.m.stuart@warwick.ac.uk

‡Korteweg-de Vries Institute for Mathematics, University of Amsterdam, The Netherlands. Email: j.h.vanzanten@uva.nl

This is the form we consider here.

We are interested in the statistical problem of recovering the drift function b given an observed path of the diffusion, $\{X_t\}_{t \in [0, T]}$, which is a solution of (1.1). Whenever application-driven insight into the form of the drift b is available, one can attempt to exploit this by postulating a parametric model for b , indexed by some finite-dimensional parameter $\theta \in \Theta \subset \mathbb{R}^d$. The statistical problem then reduces to estimating the parameter θ , see e.g. [22] for an overview of this well-researched area. In other cases however, one has to resort to nonparametric methods for making inference on the function b . Several such methods have been proposed in the literature. An incomplete list include kernel methods (e.g. [3], [22], [38]), penalized likelihood methods (e.g. [8]), and spectral approaches [2].

In this paper we investigate recently developed Bayesian methodology for estimating the drift function of a diffusion based on continuous-time observations $X^T = \{X_t\}_{t \in [0, T]}$. We consider a periodic set-up, which essentially means that we observe a diffusion on the circle. This is motivated by applications, for instance in molecular dynamics or neuroscience, in which the data consists of recordings of angles, cf. e.g. [27], [18] or [26]. We will consider Gaussian prior measures for the periodic drift function b whose inverse covariance operators are chosen from a family of even order differential operators. Recent applied work has shown that this is computationally attractive, since numerical methods for differential equations can be used for posterior sampling. Specifically, for the prior distributions we consider in this paper we will derive a weakly formulated differential equation for the posterior mean. Existing numerical methods can be used to solve this equation, allowing for posterior inference without the need to resort to Markov chain Monte Carlo methods. This numerical approach, including algorithms to accommodate both continuously and discretely observed data, is detailed in the paper [26].

In Section 2 we precisely state the inference problem of interest, and describe the properties of the family of Gaussian priors that we adopt. We postulate a prior precision operator of the form

$$\mathcal{C}_0^{-1} = \eta ((-\Delta)^p + \kappa I),$$

where Δ is the one-dimensional Laplacian, p is an integer and η, κ are real and positive hyper-parameters. Working with prior precision operators has numerous computational advantages and a central goal of this work is to develop statistical tools of analysis, in particular for posterior consistency studies, which are well-adapted to this setting. The work of [1] developed tools of analysis which do this in the context of linear inverse problems with small observational noise, and we adapt the techniques developed there to our setting.

An appealing aspect of choosing a Gaussian prior on the drift function b is conjugacy, in the sense that the posterior is Gaussian as well. Since the log-likelihood is quadratic in b (Girsanov's theorem) this is not unexpected. Formally the posterior can be computed by "completing the square". We note however that for our model, if b is distributed according to a Gaussian prior Π

and given b , the data X^T are generated by (1.1), the joint distribution of b and X^T is obviously not Gaussian in general. As a result, deriving Gaussianity of the posterior in this infinite-dimensional setting is not entirely straightforward. Section 3 is devoted to showing that, for the priors that we consider, the posterior, i.e. the conditional distribution of b given X^T , is indeed Gaussian. After a formal derivation of the associated posterior in Subsection 3.2 we rigorously prove in Theorem 3.3 below that the associated posterior is Gaussian and obtain the posterior mean and covariance structure. The posterior precision operator is again a differential operator, involving the local time of the diffusion, and the posterior mean is characterized as the unique weak solution of a $2p$ -th order differential equation. In Subsection 3.3 we outline how our Bayesian approach with Gaussian prior can be viewed as a penalized least-squares estimator, where the p th order Sobolev norm of b is penalized and the hyper-parameters η and κ quantify the degree of penalization. In the inverse problem literature this connection is known as Tikhonov-regularisation.

In Bayesian nonparametrics it is well known that careless constructions of priors can lead to inconsistent procedures and sub-optimal convergence rates (e.g. [11], [7]). Consistency or rate of convergence results are often obtained using general results that are available for various types of statistical models and that give sufficient conditions in terms of metric entropy and prior mass assumptions. See, for instance, [16], [14], [15], [36], and the references therein. In this paper however we use the explicit description of the posterior distribution, which allows us to take a rather direct approach to studying the asymptotic behaviour of our procedure. In particular, we avoid entropy or prior mass considerations.

Since the posterior involves a periodic version of the local time of the process X , the asymptotic properties of the local time play a key role in this investigation. In the present setting the existing asymptotic theory for the local time of ergodic diffusions (cf. e.g. [39], [37]) can not be used however, since we do not assume ergodicity but instead rely on the periodicity of the drift function b to accumulate information as $T \rightarrow \infty$. As a consequence, the existing posterior rate of convergence results for ergodic diffusion models of [25] do not apply. Existing limit theorems for diffusions with periodic coefficients (e.g. [4], [31], [6]) also do not suffice for our purpose. In Section 4 we therefore present new limit theorems for the local time of diffusions on the circle. These can be seen as extending and complementing the work of Bolthausen [6], who proved a uniform central limit theorem for the local time of Brownian motion on the circle (the case $b \equiv 0$ in (1.1)). For our purposes we need asymptotic tightness of the properly normalized local time in certain Sobolev spaces however, and we need the result not just for Brownian motion, but for general periodic, zero-mean drift functions b .

Having these technical tools in place we use them in combination with methods from the analysis of differential equations in Section 5 to obtain a rate of contraction result for the posterior distribution. The result states that when the true drift function b is periodic and p -regular in the Sobolev sense, then the posterior contracts around b at a rate that is essentially $T^{-(p-1/2)/(2p)}$ as $T \rightarrow \infty$.

(with respect to the L^2 -norm). In particular, we have posterior consistency.

In the concluding section we discuss several possibilities for further refinements and extensions of the present work.

2 Observation model and prior distribution

In this section we first introduce the diffusion process under study, fixing notation and describing how we exploit periodicity; see Subsection 2.1. In Subsection 2.2 we introduce the prior we place on the drift function of the diffusion, specifying the prior precision operator and collecting basic properties.

2.1 The diffusion

Consider the stochastic differential equation (SDE)

$$dX_t = b(X_t) dt + dW_t, \quad X_0 = 0, \quad (2.1)$$

where W is standard Brownian motion and $b : \mathbb{R} \rightarrow \mathbb{R}$ is a continuously differentiable, 1-periodic drift function with zero mean, i.e. $b(x+k) = b(x)$ for all $x \in \mathbb{R}$ and $k \in \mathbb{Z}$ and $\int_0^1 b(x) dx = 0$. We let \mathbb{T} denote the circle $[0, 1)$ so that we can also write $b : \mathbb{T} \rightarrow \mathbb{R}$ and we summarize the assumptions on b by writing $b \in \dot{C}^1(\mathbb{T})$, the dot denoting mean zero.

We assume the mean zero property of b for technical reasons. From the perspective of the statistical problem of nonparametrically estimating the drift b it is not a serious restriction. Note that if the diffusion X has a periodic drift function b with mean $\bar{b} = \int_0^1 b(x) dx$, then the process $\{X_t - \bar{b}t\}_{t \geq 0}$ has the zero-mean drift function $b - \bar{b}$. In practice, the mean can be removed in a preliminary step using an auxiliary estimator for \bar{b} . The simple estimator X_T/T can be used for instance. It converges in probability to \bar{b} at the rate $T^{-1/2}$ as $T \rightarrow \infty$ (cf. [4], Theorem 3), which is faster than the rates we obtain for the nonparametric problem of estimating the centred drift function.

For every $b \in \dot{C}^1(\mathbb{T})$ the SDE (2.1) has a unique weak solution (see e.g. Theorems 6.1.6 and 6.2.1 in [12], p. 214). For $T > 0$, we denote the law that this solution generates on the canonical path space $C[0, T]$ by \mathbb{P}_b^T . In particular \mathbb{P}_0^T is the Wiener measure on $C[0, T]$. By Girsanov's theorem the laws \mathbb{P}_b^T , $b \in \dot{C}^1(\mathbb{T})$, are all equivalent on $C[0, T]$. If two measurable maps of X are almost surely (a.s.) equal under some $\mathbb{P}_{b_0}^T$ they are therefore a.s. equal under any of the laws \mathbb{P}_b^T , and we will simply write that they are equal a.s..

We drop the superscript T and denote the sample path of (2.1) by $X \in C[0, T]$. The Radon-Nikodym derivative of \mathbb{P}_b^T relative to the Wiener measure satisfies

$$\frac{d\mathbb{P}_b^T}{d\mathbb{P}_0^T}(X) = \exp\left(-\frac{1}{2} \int_0^T b^2(X_t) dt + \int_0^T b(X_t) dX_t\right)$$

almost surely, by Girsanov's theorem (e.g. [24]). Observe that by Itô's formula the likelihood can be rewritten as

$$\frac{d\mathbb{P}_b^T}{d\mathbb{P}_0^T}(X) = \exp\left(-\Phi_T(b; X)\right)$$

a.s., where

$$\Phi_T(b; X) = \frac{1}{2} \int_0^T (b^2(X_t) + b'(X_t)) dt + B(X_0) - B(X_T) \quad (2.2)$$

and $B' = b$. Note that B is also 1-periodic, since b has average zero.

It will be convenient to write the integrals in the expression for Φ_T in terms of the local time of the process X . Let $(L_t(x; X) : t \geq 0, x \in \mathbb{R})$ be the semi-martingale local time of X , so that

$$\int_{-\infty}^{\infty} f(x) L_T(x; X) dx = \int_0^T f(X_s) ds \quad (2.3)$$

holds a.s. for any bounded, measurable $f : \mathbb{R} \rightarrow \mathbb{R}$. Defining also the random variables $\chi_T(x; X)$ by

$$\chi_T(x; X) = \begin{cases} 1 & \text{if } X_0 < x < X_T, \\ -1 & \text{if } X_T < x < X_0, \\ 0 & \text{otherwise,} \end{cases}$$

we may then write

$$\Phi_T(b; X) = \frac{1}{2} \int_{\mathbb{R}} \left(L_T(x; X)(b^2(x) + b'(x)) - 2\chi_T(x; X)b(x) \right) dx. \quad (2.4)$$

In view of the periodicity of the functions involved it is sensible to introduce a periodic version L° of the local time L by defining

$$L_T^\circ(x; X) = \sum_{k \in \mathbb{Z}} L_T(x + k; X)$$

for $x \in \mathbb{T}$. Note that for every $T > 0$, the random function $x \mapsto L_T(x; X)$ a.s. is a continuous function with support included in the compact interval $[\min_{t \leq T} X_t, \max_{t \leq T} X_t]$. Hence the infinite sum can actually be restricted to the finitely many integers in the interval $[\min_{t \leq T} X_t - 1, \max_{t \leq T} X_t + 1]$. Hence the sum is well defined and $x \mapsto L_T^\circ(x; X)$ is a continuous random function on \mathbb{T} . In particular, we have that the norms $\|L_T^\circ(\cdot; X)\|_\infty$ and $\|L_T^\circ(\cdot; X)\|_{L^2}$ are a.s. finite.

It follows from (2.3) that for any 1-periodic, bounded, measurable function f and $T \geq 0$,

$$\int_0^T f(X_u) du = \int_0^1 f(x) L_T^\circ(x; X) dx. \quad (2.5)$$

Exploiting the periodicity of b and B and introducing the corresponding periodized version $\chi_T^\circ(\cdot; X)$ of $\chi_T(\cdot; X)$, we can then rewrite (2.4) as

$$\Phi_T(b; X) = \frac{1}{2} \int_0^1 \left(L_T^\circ(x; X)(b^2(x) + b'(x)) - 2\chi_T^\circ(x; X)b(x) \right) dx. \quad (2.6)$$

Summarizing, we have the following lemma.

Lemma 2.1. *For every $b \in \dot{C}^1(\mathbb{T})$ and $T > 0$ the law \mathbb{P}_b^T is equivalent to \mathbb{P}_0^T on $C[0, T]$ and*

$$\frac{d\mathbb{P}_b^T}{d\mathbb{P}_0^T}(X) = \exp\left(-\Phi_T(b; X)\right),$$

a.s., where Φ_T is given by (2.6).

2.2 The prior

We will assume that we observe a solution of the SDE (2.1) up to time $T > 0$, for some $b \in \dot{C}^1(\mathbb{T})$. To make inference on b we endow it with a centred Gaussian prior Π . We will view the prior as a centred Gaussian measure on $L^2(\mathbb{T})$ and define it through its covariance operator \mathcal{C}_0 , or, rather, through its precision operator \mathcal{C}_0^{-1} . Specifically, we fix hyper-parameters $\eta, \kappa > 0$ and $p \in \{2, 3, \dots\}$ and consider the operator \mathcal{C}_0 with densely defined inverse

$$\mathcal{C}_0^{-1} = \eta((-\Delta)^p + \kappa I), \quad (2.7)$$

where Δ denotes the one-dimensional Laplacian, I is the identity and the domain of \mathcal{C}_0^{-1} is given by $D(\mathcal{C}_0^{-1}) = \dot{H}^{2p}(\mathbb{T})$, the space of mean-zero functions in the Sobolev space $H^{2p}(\mathbb{T})$ of functions in $L^2(\mathbb{T})$ with $2p$ square integrable weak derivatives.

To see that \mathcal{C}_0 is indeed a valid covariance operator and hence the prior is well defined, consider the orthonormal basis ϕ_k of $\dot{L}^2(\mathbb{T})$, which is by definition the space of mean-zero functions in $L^2(\mathbb{T})$, given by

$$\begin{aligned} \phi_{2k}(x) &= \sqrt{2} \cos(2\pi kx), \\ \phi_{2k-1}(x) &= \sqrt{2} \sin(2\pi kx), \end{aligned}$$

for $k \in \mathbb{N}$. The functions ϕ_k belong to the domain $\dot{H}^{2p}(\mathbb{T})$ of the operator (2.7) and

$$\begin{aligned} \mathcal{C}_0^{-1} \phi_{2k} &= (\eta(4\pi^2 k^2)^p + \eta\kappa) \phi_{2k}, \\ \mathcal{C}_0^{-1} \phi_{2k-1} &= (\eta(4\pi^2 k^2)^p + \eta\kappa) \phi_{2k-1}, \end{aligned}$$

for $k \in \mathbb{N}$. It follows that \mathcal{C}_0 is the operator on $\dot{L}^2(\mathbb{T})$ which is diagonalized by the basis ϕ_k , with eigenvalues

$$\lambda_k = \left(\eta \left(4\pi^2 \left\lceil \frac{k}{2} \right\rceil^2 \right)^p + \eta\kappa \right)^{-1}. \quad (2.8)$$

Thus \mathcal{C}_0 is positive definite, symmetric, and trace-class and hence a covariance operator on $\dot{L}^2(\mathbb{T})$. (It extends to a covariance operator on the whole space $L^2(\mathbb{T})$ by setting $\mathcal{C}_0 1 = 0$.)

The integer p in (2.7) controls the regularity of the prior Π and we assume $p \geq 2$ to ensure that the drift is C^1 (see Lemma 2.2 below). The parameter $\eta > 0$ sets an overall scale for the precision. The parameter κ allows us to shift the precisions in every mode by a uniform amount. We employ $\kappa > 0$ as it simplifies some of the analysis, but $\kappa = 0$ could be included in the analysis with further work. Likewise we have assumed a mean zero prior, but extensions to include a mean could be made.

The preceding calculations show that the prior Π is the law of the centred Gaussian process $V = \{V_x\}_{x \in \mathbb{T}}$ defined by

$$V_x = \sum_{k \in \mathbb{N}} \sqrt{\lambda_k} \phi_k(x) Z_k, \quad (2.9)$$

for Z_1, Z_2, \dots independent, standard Gaussian random variables. Using this series representation a number of basic properties of the prior can easily be derived.

Lemma 2.2.

- (i) *There exists a version of V which a.s. has sample paths that are Hölder continuous of order α , for every $\alpha < p - 1/2$.*
- (ii) *The reproducing kernel Hilbert space of V is the Sobolev space $\dot{H}^p(\mathbb{T})$.*
- (iii) *The L^2 -support of Π is $\dot{L}^2(\mathbb{T})$.*

Proof. For the first statement, note that $\sqrt{\lambda_k} \sim k^{-p}$ asymptotically. Using also the differential relations between the basis functions ϕ_k it is straightforward to see that the process V has $p - 1 \geq 1$ weak derivatives in the L^2 -sense. Moreover, using Kolmogorov's classical continuity theorem it can be shown that this $(p - 1)$ st derivative has a version with sample paths that are Hölder continuous of order γ for every $\gamma < 1/2$. Combining this we see that V has a version with α -Hölder sample paths, for every $\alpha < p - 1/2$. In particular, it holds that all the mass of the prior Π is concentrated on $\dot{C}^1(\mathbb{T})$.

The Karhunen-Loève expansion (2.9) shows that the reproducing kernel Hilbert space (RKHS) of the prior is given by $\mathbb{H} = \{\sum_{k \geq 1} c_k \phi_k : \sum c_k^2 / \lambda_k < \infty\}$ (see for instance [41], Theorem 4.1). Since $1/\lambda_k \sim k^{2p}$ this implies that $\mathbb{H} = \dot{H}^p(\mathbb{T})$, proving the second statement.

The final statement follows from the second one, since the L^2 -support is the L^2 -closure of \mathbb{H} ([41], Lemma 5.1). \square

Note in particular that the lemma shows that we can view the prior Π as a Gaussian measure on any of the separable Banach spaces $L^2(\mathbb{T})$, $C(\mathbb{T})$, $C^k(\mathbb{T})$ or $H^k(\mathbb{T})$, for $k \leq p - 1$.

3 Posterior distribution

3.1 Bayes' formula

We recall that X denotes the path $\{X_t\}_{t \in [0, T]}$. If we endow $C^1(\mathbb{T})$ with its Hölder norm and $C[0, T]$ with the uniform norm, then expression (2.2) shows that the negative log-likelihood $(b, x) \mapsto \Phi_T(b; x)$ (has a version that) is Borel-measurable as a map from $C^1(\mathbb{T}) \times C[0, T] \rightarrow \mathbb{R}$. Since we can view Π as a measure on $C^1(\mathbb{T})$, it follows that we have a well-defined Borel measure $\Pi(db)\mathbb{P}_b^T(dx)$ on $C^1(\mathbb{T}) \times C[0, T]$, which is the joint law of b and X in the Bayesian set-up

$$b \sim \Pi, \quad X | b \sim (2.1).$$

The posterior distribution, i.e. the conditional distribution of b given X , is then well-defined as well and given by

$$\begin{aligned} \Pi(B | X) &= \frac{1}{Z} \int_B \exp(-\Phi_T(b; X)) \Pi(db), \\ Z &= \int_{C^1(\mathbb{T})} \exp(-\Phi_T(b; X)) \Pi(db), \end{aligned} \tag{3.1}$$

Lemma 3.1. *The random Borel measure $B \mapsto \Pi(B | X)$ on $C^1(\mathbb{T})$ given by (3.1) is a.s. well defined.*

Proof. By Lemma 5.3 of [17], the posterior is well-defined if $Z > 0$ a.s.. To see that the latter condition is fulfilled, observe that

$$|\Phi_T(b; X)| \lesssim (1 + \|L_T^\circ(\cdot; X)\|_\infty)(\|b\|_\infty^2 + \|b'\|_\infty).$$

Since Π is a centred Gaussian distribution on the separable Banach space $C^1(\mathbb{T})$, endowed with its Hölder norm, we have $\Pi(b : \|b\|_\infty + \|b'\|_\infty < \infty) = 1$. Together this gives the a.s. positivity of Z , since $\|L_T^\circ(\cdot; X)\|_\infty < \infty$ a.s. (Here, and elsewhere, $a \lesssim b$ means that a is less than an irrelevant constant times b .) \square

We have now defined the posterior as a measure on $C^1(\mathbb{T})$, but since the prior is in fact a probability measure on $C^\alpha(\mathbb{T})$ for every $\alpha < p - 1/2$ (see the preceding section), it is a Borel measure on these Hölder spaces as well. We can of course also view it as a measure on $C(\mathbb{T})$ or $L^2(\mathbb{T})$.

3.2 Formal computation of the posterior

The next goal is to characterize the posterior. We proceed first strictly formally and non-rigorously. Very loosely speaking, we have that the prior Π has a “density” proportional to

$$b \mapsto \exp\left(-\frac{1}{2} \int_0^1 b(x) \mathcal{C}_0^{-1} b(x) dx\right) \tag{3.2}$$

and the negative log-likelihood also has a quadratic form, given by (2.6). This suggests that the posterior is again Gaussian. Formally completing the square gives the relations

$$\mathcal{C}_T^{-1} = \mathcal{C}_0^{-1} + L_T^\circ(\cdot; X)I, \quad (3.3)$$

$$\mathcal{C}_T^{-1}\hat{b}_T = \frac{1}{2}(L_T^\circ(\cdot; X))' + \chi_T^\circ(\cdot; X) \quad (3.4)$$

for the posterior mean \hat{b}_T and the posterior precision operator \mathcal{C}_T^{-1} .

As detailed in the preceding section we assume that the prior covariance operator is given by (2.7), with integer $p \geq 2$, $\eta, \kappa > 0$, Δ the one-dimensional Laplacian and $D(\mathcal{C}_0^{-1}) = \dot{H}^{2p}(\mathbb{T})$. In that case (3.3) gives

$$\mathcal{C}_T^{-1} = \eta(-\Delta)^p + (\eta\kappa + L_T^\circ(\cdot; X))I \quad (3.5)$$

and $D(\mathcal{C}_T^{-1}) = \dot{H}^{2p}(\mathbb{T})$. By standard application of the Lax-Milgram theory (see [13], Section 6.2), it follows that the equation $\mathcal{C}_T^{-1}f = g$ has a unique weak solution in $\dot{H}^p(\mathbb{T})$ for every $g \in \dot{H}^{-p}(\mathbb{T})$; see [29], Appendix A, for definition and properties of the Sobolev spaces $\dot{H}^{2p}(\mathbb{T})$. From this it follows that \mathcal{C}_T is well defined on all of $\dot{H}^{-p}(\mathbb{T})$. Moreover, \mathcal{C}_T is a bounded operator from $\dot{H}^{-p}(\mathbb{T})$ into $\dot{H}^p(\mathbb{T})$, since \mathcal{C}_T^{-1} is coercive. If $g \in L^2(\mathbb{T})$ then the weak solution is more regular and, in fact, lies in $\dot{H}^{2p}(\mathbb{T})$; see [13], Section 6.3.

The ordinary derivative of local time is not defined, and indeed is not an element of $L^2(\mathbb{T})$. Thus we will have to interpret (3.4) in a weak sense. In order to enable us to do this, in Subsection 3.4 we consider the variational formulation of equation (3.4). As a precursor to this, the next subsection is devoted to observing that the differential equation for the mean arises as the Euler-Lagrange equation for a certain variational problem, yielding an interesting connection with penalized least-squares estimation.

3.3 Connection with penalized least squares

Here we demonstrate the fact that the posterior mean \hat{b}_T given by (3.4) can be viewed as a penalized least-squares estimator in the case $p = 2$. Formally, the SDE (2.1) can be written as

$$\dot{X}_t = b(X_t) + \dot{W}_t,$$

where the dot denotes differentiation with respect to t (obviously, the derivatives \dot{X} and \dot{W} do not exist in the ordinary sense.) This is just a continuous-time version of a standard nonparametric regression model and for a drift function u , we can view the integral

$$\int_0^T (\dot{X}_t - u(X_t))^2 dt$$

as a residual sum of squares. A penalized least-squares procedure consists in adding a penalty term to this quantity and minimizing the resulting criterion

over u . Expanding the square in the preceding integral shows that this is equivalent to minimizing

$$u \mapsto - \int_0^T u(X_t) dX_t + \frac{1}{2} \int_0^T u^2(X_t) dt + P(u),$$

over an appropriate space of functions, where $P(u)$ is the penalty.

If the function u is smooth and periodic, then by Itô's formula and the definitions of L° and χ° , we have, with U a primitive function of u ,

$$\begin{aligned} \int_0^T u(X_t) dX_t &= U(X_T) - U(X_0) - \frac{1}{2} \int_0^T u'(X_t) dt \\ &= \int_0^1 u(x) \chi_T^\circ(x; X) dx - \frac{1}{2} \int_0^1 L_T^\circ(x; X) u'(x) dx \end{aligned}$$

and

$$\int_0^T u^2(X_t) dt = \int_0^1 u^2(x) L_T^\circ(x; X) dx.$$

Hence, if the functions u over which the minimization takes place are smooth enough, the criterion can also be written as

$$u \mapsto \int_0^1 \left(\frac{1}{2} u^2(x) L_T^\circ(x; X) + \frac{1}{2} u'(x) L_T^\circ(x; X) - u(x) \chi_T^\circ(x; X) \right) dx + P(u).$$

Now consider a Sobolev-type penalty term of the form

$$P(u) = \frac{1}{2} \eta \left(\kappa \int_0^1 (u(x))^2 dx + \int_0^1 (u''(x))^2 dx \right),$$

for constants $\eta, \kappa > 0$. Then the objective functional $u \mapsto \Lambda(u; X)$ takes the form

$$\Lambda(u; X) = \int_0^1 \left(\frac{1}{2} u^2(\eta\kappa + L_T^\circ(X)) + \frac{1}{2} u' L_T^\circ(X) - u \chi_T^\circ(X) + \frac{1}{2} \eta (u'')^2 \right) dx,$$

where we omitted explicit dependence on x to lighten notation. To minimize this functional, simply take its variational derivative in the direction v , i.e. compute the limit $\lim_{\epsilon \rightarrow 0} (\Lambda(u + \epsilon v; X) - \Lambda(u; X)) / \epsilon$, for a smooth test function v :

$$\frac{\delta \Lambda}{\delta u}(v) = \int_0^1 \left(uv L_T^\circ(X) - \frac{1}{2} v (L_T^\circ)'(X) - v \chi_T^\circ(X) + \eta v'' u'' + \eta \kappa uv \right) dx$$

A further integration by parts (where the boundary terms vanish due to periodicity) now yields the form

$$\frac{\delta \Lambda}{\delta u}(v) = \int_0^1 v \left(u L_T^\circ(X) - \frac{1}{2} (L_T^\circ)'(X) - \chi_T^\circ(X) + \eta u^{(4)} + \eta \kappa u \right) dx$$

from which it is evident that equating the variational derivative to zero for all smooth test functions yields exactly the posterior mean obtained in (3.4) for the case $p = 2$:

$$\eta u^{(4)} + (\eta\kappa + L_T^\circ(X))u = \frac{1}{2}(L_T^\circ)'(X) + \chi_T^\circ(X).$$

In the context of inverse problems, adding the square of the norm of the underlying vector space is known as (generalized) Tikhonov regularization, and the connection to Bayesian inference with a Gaussian prior is well established in general, see [34]. It may be viewed as a natural extension of the approach of Wahba [42] from regression to the diffusion process setting. The case of regularization through higher order derivatives in the penalization term P is similar.

3.4 Weak variational formulation for the posterior mean

In the preceding section we remarked that the RKHS of the Gaussian prior equals the Sobolev space $\dot{H}^p(\mathbb{T})$. Below we prove that the posterior is a.s. a Gaussian measure. Moreover, since the denominator Z in (3.1) is positive a.s., the posterior is equivalent to the prior. It follows that the posterior mean \hat{b}_T is a.s. an element of $\dot{H}^p(\mathbb{T})$. By saying it is a weak solution to (3.4) we mean that it solves the following weak form of the associated variational principle:

$$a(\hat{b}_T, v; X) = r(v; X) \quad \text{for every } v \in \dot{H}^p(\mathbb{T}), \quad (3.6)$$

where the bilinear form $a(\cdot, \cdot; X) : \dot{H}^p(\mathbb{T}) \times \dot{H}^p(\mathbb{T}) \rightarrow \mathbb{R}$ and the linear form $r(\cdot; X) : \dot{H}^p(\mathbb{T}) \rightarrow \mathbb{R}$ are defined by

$$\begin{aligned} a(u, v; X) &= \eta \int u^{(p)}(x)v^{(p)}(x) dx + \eta\kappa \int u(x)v(x) dx + \int u(x)v(x)L_T^\circ(x; X) dx, \\ r(v; X) &= -\frac{1}{2} \int v'(x)L_T^\circ(x; X) dx + \int v(x)\chi_T^\circ(x; X) dx. \end{aligned}$$

The following lemma records the essential properties of a and r and the associated variational problem.

Lemma 3.2. *The following statements hold almost surely:*

(i) $a(\cdot, \cdot; X)$ is bilinear, symmetric, continuous and coercive:

$$a(v_1, v_2; X) \leq (\eta + \eta\kappa + \|L_T^\circ(\cdot; X)\|_\infty) \|v_1\|_{H^p} \|v_2\|_{H^p}$$

for $v_1, v_2 \in \dot{H}^p(\mathbb{T})$ and for some constant $c > 0$, $a(v, v; X) \geq c \|v\|_{H^p}^2$ for all $v \in \dot{H}^p(\mathbb{T})$.

(ii) $r(\cdot; X)$ is linear and bounded:

$$|r(v; X)| \leq \left(\frac{1}{2} \|L_T^\circ(\cdot; X)\|_{L^2} + \|\chi_T^\circ(\cdot; X)\|_{L^2} \right) \|v\|_{H^p}$$

for all $v \in \dot{H}^p(\mathbb{T})$.

(iii) There exists a unique $u \in \dot{H}^p(\mathbb{T})$ such that $a(u, v; X) = r(v; X)$ for all $v \in \dot{H}^p(\mathbb{T})$.

Proof. (i) Bi-linearity, symmetry and continuity follow straightforwardly from the definition of a . Coercivity follows easily from the positivity of η and κ and the Poincaré inequality (see [29], Proposition 5.8.) (ii) Again, straightforward. (iii) Follows from (i) and (ii) by the Lax-Milgram Lemma, see [13], Section 6.2. \square

3.5 Characterization of the posterior

We can now prove that the posterior is Gaussian and characterize its mean and covariance operator. Recall that by saying that \hat{b}_T is a weak solution of the differential equation (3.4) we mean that it solves the variational problem (3.6).

Theorem 3.3. *Almost surely, the posterior $\Pi(\cdot | X)$ is a Gaussian measure on $L^2(\mathbb{T})$. Its covariance operator \mathcal{C}_T is given by (3.5) and its mean \hat{b}_T is the unique weak solution of (3.4).*

Proof. For $n \in \mathbb{N}$, let $P_n : L^2(\mathbb{T}) \rightarrow L^2(\mathbb{T})$ be the orthogonal projection onto the linear span V_n of the first n basis functions ϕ_1, \dots, ϕ_n . Let the random measure $\Pi_n(\cdot | X)$ be given by

$$\begin{aligned} \Pi_n(B | X) &= \frac{1}{Z_n} \int_B \exp(-\Phi_T(P_n b; X)) \Pi(db), \\ Z_n &= \int_{C^1(\mathbb{T})} \exp(-\Phi_T(P_n b; X)) \Pi(db), \end{aligned}$$

for Borel sets $B \subset C^1(\mathbb{T})$. The fact that this random measure is well defined can be argued exactly as in Section 3.1.

For $b \in \dot{C}^1(\mathbb{T})$ it holds that $P_n b \rightarrow b$ in $H^1(\mathbb{T})$ as $n \rightarrow \infty$. It is easily seen from (2.6) that the random map $b \mapsto \Phi_T(b; X)$ is a.s. $H^1(\mathbb{T})$ -continuous. It follows that a.s., $b \mapsto \Phi_T(P_n b; X)$ converges point-wise to $\Phi_T(\cdot; X)$ on $\dot{C}^1(\mathbb{T})$. By Lemma 3.4 below, there exists for every $\varepsilon > 0$ a constant $K(\varepsilon)$ such that

$$-\Phi_n(b; X) \leq \varepsilon \|b\|_{H^1}^2 + K(\varepsilon)(1 + \|L_T^\circ(\cdot; X)\|_{L^2}^2),$$

and hence

$$e^{-\Phi_n(P_n b; X)} \leq e^{K(\varepsilon)(1 + \|L_T^\circ(\cdot; X)\|_{L^2}^2)} e^{\varepsilon \|b\|_{H^1}^2}.$$

Since Π can be viewed as a Gaussian measure on $H^1(\mathbb{T})$, Fernique's theorem implies that a.s., the right-hand side of the last display is a Π -integrable function of b for $\varepsilon > 0$ small enough (see [5], Theorem 2.8.5). Hence, by dominated convergence, we can conclude that $Z_n \rightarrow Z$ almost surely. The same reasoning shows that for every Borel set $B \subset C^1(\mathbb{T})$, it a.s. holds that

$$\int_B \exp(-\Phi_T(P_n b; X)) \Pi(db) \rightarrow \int_B \exp(-\Phi_T(b; X)) \Pi(db)$$

as $n \rightarrow \infty$, where we rewrite the integral as an integral over $C^1(\mathbb{T})$ and then exploit boundedness of the indicator function $\chi_B(\cdot)$ thus introduced into the integrand.

Hence, we have that with probability 1, the measures $\Pi_n(\cdot | X)$ converge weakly to the posterior $\Pi(\cdot | X)$. Note that the weak convergence takes place in $C^1(\mathbb{T})$, but then in $L^2(\mathbb{T})$ as well. Since the measures $\Pi_n(\cdot | X)$ are easily seen to be Gaussian, the measure $\Pi(\cdot | X)$ must be Gaussian as well.

If we view $\dot{L}^2(\mathbb{T})$ as the product of V_n and V_n^\perp , then by construction the measure $\Pi_n(\cdot | X)$ is a product of Gaussian measures on V_n and V_n^\perp . The measure on V_n really has density proportional to (3.2), relative to the push-forward measure of the Lebesgue measure on \mathbb{R}^n under the map $(c_1, \dots, c_n) \mapsto \sum c_k \phi_k$. The formal arguments given in Subsection 3.2 can therefore be made rigorous, showing that this factor is a Gaussian measure on V_n with covariance operator $P_n \mathcal{C}_T P_n$ and mean $b_n \in V_n$ which solves the variational problem

$$a(b_n, v; X) = r(v; X)$$

for every $v \in V_n$. The measure on V_n^\perp has mean zero, so b_n is in fact the mean of the whole measure $\Pi_n(\cdot | X)$. The covariance operator of the measure on V_n^\perp is given by $(I - P_n) \mathcal{C}_0 (I - P_n)$.

Next we prove that the posterior mean \hat{b}_T is the weak solution of (3.4). By Lemma 3.2 there a.s. exists a unique $u \in \dot{H}^p(\mathbb{T})$ such that $a(u, v; X) = r(v; X)$ for all $v \in \dot{H}^p(\mathbb{T})$. Standard Galerkin method arguments show that for the mean of $\Pi_n(\cdot | X)$ we have $b_n \rightarrow u$ in $\dot{H}^p(\mathbb{T})$. Indeed, let $e_n = u - b_n$. Then we have the orthogonality property $a(e_n, v; X) = 0$ for all $v \in V_n$. Using the continuity and coercivity of $a(\cdot, \cdot; X)$, cf. Lemma 3.2, it follows that for $v \in V_n$,

$$\begin{aligned} c \|e_n\|_{H^p}^2 &\leq a(e_n, e_n; X) \\ &= a(e_n, u - v; X) \\ &\leq (\eta + \eta\kappa + \|L_T^\circ(\cdot; X)\|_\infty) \|e_n\|_{H^p} \|u - v\|_{H^p}. \end{aligned}$$

Hence, for every $v \in V_n$ we have

$$c \|e_n\|_{H^p} \leq (\eta + \eta\kappa + \|L_T^\circ(\cdot; X)\|_\infty) \|u - v\|_{H^p}.$$

By taking $v = P_n u$ we then see that $b_n \rightarrow u$ in $H^p(\mathbb{T})$. On the other hand, by the weak convergence found above, b_n converges a.s. to the posterior mean \hat{b}_T in $L^2(\mathbb{T})$ (see [5], Example 3.8.15). We conclude that \hat{b}_T a.s. equals the unique weak solution u of (3.4), as required.

It remains to show that the covariance operator of the posterior is given by (3.5). Let $\Sigma_n = P_n \mathcal{C}_T P_n + (I - P_n) \mathcal{C}_0 (I - P_n)$ be the covariance operator of $\Pi_n(\cdot | X)$ and let Σ be the covariance operator of the posterior $\Pi(\cdot | X)$. Since the measures converge weakly and are Gaussian, we have that for every $f \in L^2(\mathbb{T})$, $\Sigma_n f \rightarrow \Sigma f$ in $L^2(\mathbb{T})$ (cf. Example 3.8.15 of [5] again). On the other hand, for $n > k$ and $g \in \dot{L}^2(\mathbb{T})$ we have

$$\begin{aligned} \left| \langle g, \Sigma_n \phi_k \rangle - \langle g, \mathcal{C}_T \phi_k \rangle_{L^2} \right| &= \left| \langle g, (P_n - I) \mathcal{C}_T \phi_k \rangle_{L^2} \right| \\ &\leq \|(P_n - I)g\|_{L^2} \|\mathcal{C}_T \phi_k\|_{L^2} \rightarrow 0, \end{aligned}$$

hence $\Sigma_n \phi_k$ converges weakly to $\mathcal{C}_T \phi_k$. It follows that $\Sigma \phi_k = \mathcal{C}_T \phi_k$ for every k and the proof is complete. \square

Lemma 3.4. *For every $\varepsilon > 0$ there exists a constant $K(\varepsilon) > 0$ such that*

$$-\Phi(b; X) \leq \varepsilon \|b\|_{H^1}^2 + K(\varepsilon)(1 + \|L_T^\circ(\cdot; X)\|_{L^2}^2).$$

Proof. It follows from (2.6) that

$$-\Phi(b; X) \leq +\frac{1}{2} \int_0^1 L_T^\circ(x; X) |b'(x)| dx + \int_0^1 |b(x)| dx.$$

Now note that for every $\beta > 0$ and $f, g \in L^2(\mathbb{T})$, it holds that $2\langle f, g \rangle \leq \beta \|f\|^2 + \beta^{-1} \|g\|^2$ (“Young’s inequality with ε ”, with $p = q = 2$). Applying this to both integrals on the right we get

$$\begin{aligned} -\Phi(b; X) &\leq \frac{\beta}{4} \|L_T^\circ(\cdot; X)\|_{L^2}^2 + \frac{1}{4\beta} \|b'\|_{L^2}^2 + \frac{\beta}{2} + \frac{1}{2\beta} \|b\|_{L^2}^2 \\ &\leq \varepsilon \|b\|_{H^1}^2 + \frac{1}{4\varepsilon} + \frac{1}{8\varepsilon} \|L_T^\circ(\cdot; X)\|_{L^2}^2, \end{aligned}$$

where $\varepsilon = (2\beta)^{-1}$, so $K(\varepsilon) = (4\varepsilon)^{-1}$. \square

4 Asymptotic behaviour of the local time

In the next section we will investigate the asymptotic behaviour of the posterior, using the characterization provided by Theorem 3.3. Since (3.3) and (3.4) involve the periodic local time $L_T^\circ(\cdot; X)$, the asymptotic properties of that random function play a key role.

The results we establish in this section can be seen as complementing and extending the work of Bolthausen [6] in which it is proved that if X is Brownian motion (i.e. $b \equiv 0$ in (2.1)), then the random functions

$$\sqrt{T} \left(\frac{1}{T} L_T^\circ(\cdot; X) - 1 \right)$$

converge weakly in the space $C(\mathbb{T})$ to a Gaussian random map as $T \rightarrow \infty$. For our purposes we need a similar result in the Sobolev space $H^\alpha(\mathbb{T})$, for $\alpha < 1/2$, and we need the result not just for Brownian motion, but for general periodic, zero-mean drift functions b . In fact asymptotic tightness, rather than weak convergence, suffices for our purposes and it is this which we prove. In addition, we need the associated uniform law of large numbers which states that

$$\frac{1}{T} L_T^\circ(\cdot; X)$$

converges uniformly as $T \rightarrow \infty$. Similar statements were obtained for ergodic diffusions in the papers [39] and [37]. In the present periodic setting however, completely different arguments are necessary.

Given $b \in \dot{C}(\mathbb{T})$ we define the probability density ρ on $[0, 1]$ by

$$\rho(x) = C \exp\left(2 \int_0^x b(y) dy\right), \quad x \in [0, 1], \quad (4.1)$$

where $C > 0$ is the normalization constant that ensures that ρ integrates to 1. In the one-dimensional diffusion language, ρ is the restriction to $[0, 1]$ of the speed density of the diffusion, normalized so that it becomes a probability density. Note that since b has mean zero, ρ satisfies $\rho(0) = \rho(1)$ and enjoys a natural extension to a periodic function.

We use the standard notation that $Y_T = O_{\mathbb{P}}(a_T)$ for the family of random variables $\{Y_T\}$ and the deterministic family of positive numbers $\{a_T\}$ if the family $\{Y_T/a_T\}$ is asymptotically tight as $T \rightarrow \infty$ with respect to the probability space underlying the random variables $\{Y_T\}$.

Theorem 4.1.

(i) *It almost surely holds that*

$$\sup_{x \in \mathbb{T}} \left| \frac{1}{T} L_T^\circ(x; X) - \rho(x) \right| \rightarrow 0$$

as $T \rightarrow \infty$.

(ii) *For every $\alpha < 1/2$, the random maps*

$$x \mapsto \sqrt{T} \left(\frac{1}{T} L_T^\circ(x; X) - \rho(x) \right)$$

are asymptotically tight in $H^\alpha(\mathbb{T})$ as $T \rightarrow \infty$. In particular, for every $\alpha < 1/2$,

$$\left\| \frac{1}{T} L_T^\circ(\cdot; X) - \rho \right\|_{H^\alpha} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right)$$

as $T \rightarrow \infty$.

The proof of the theorem is long and therefore deferred to Section 6 in order to keep the overarching arguments in this paper, which are aimed to proving posterior consistency, to the fore. In the following subsection about posterior contraction rates we need the fact that $X_T = O_{\mathbb{P}}(\sqrt{T})$ for the diffusions with periodic drift under consideration. This follows for instance from the results of [4], but we can alternatively derive it from the preceding theorem.

Corollary 4.2. *For every $b \in \dot{C}(\mathbb{T})$, the weak solution $X = (X_t : t \geq 0)$ of the SDE (2.1) satisfies $X_T = O_{\mathbb{P}}(\sqrt{T})$ as $T \rightarrow \infty$.*

Proof. We have

$$X_T = \int_0^T b(X_s) ds + W_T$$

for a standard Brownian motion W . Since b is 1-periodic, the integral can be rewritten in terms of the periodic local time L° . Moreover, (4.1) implies that ρ is 1-periodic as well and $\rho' = 2b\rho$, hence

$$\int_0^1 b(x)\rho(x) dx = \frac{1}{2}(\rho(1) - \rho(0)) = 0.$$

It follows that

$$\begin{aligned} |X_T| &\leq T \left| \int_0^1 b(x) \left(\frac{L_T^\circ(x; X)}{T} - \rho(x) \right) dx \right| + |W_T| \\ &\leq T \|b\|_{L^2} \left\| \frac{1}{T} L_T^\circ(\cdot; X) - \rho \right\|_{L^2} + |W_T|. \end{aligned}$$

By the preceding theorem, this is $O_{\mathbb{P}}(\sqrt{T})$. \square

Note that statement (i) of Theorem 4.1 implies that for an integrable, 1-periodic function f , we have the strong law of large numbers

$$\frac{1}{T} \int_0^T f(X_t) dt \rightarrow \int_0^1 f(x)\rho(x) dx$$

a.s. as $T \rightarrow \infty$. Moreover, if f is also square integrable, statement (ii) implies that

$$\frac{1}{T} \int_0^T f(X_t) dt - \int_0^1 f(x)\rho(x) dx = O_{\mathbb{P}}\left(\frac{1}{\sqrt{T}}\right).$$

Result of this type can be found also in [31] and are of independent interest. They can for instance be useful in the asymptotic analysis of other statistical procedures for the periodic diffusion models we are considering. Uniform Glivenko-Cantelli and Donsker-type statements could be derived using our approach as well, similar to the results for ergodic diffusions in [38] and [39]. Since this is outside the scope of the present paper however, we do not elaborate on this any further here.

5 Posterior contraction rates

In this section we use the characterization of the posterior provided by Theorem 3.3 and the asymptotic behaviour of the local time established in Theorem 4.1 to study the rate at which the posterior contracts around the true drift function, which we denote by b_0 to emphasize that the results are frequentist in nature. In particular \mathbb{P}_{b_0} denotes the underlying probability measure corresponding to the true drift function b_0 , and the notation $O_{\mathbb{P}_{b_0}}$ refers to asymptotic tightness under this measure.

The first theorem concerns the rate of convergence of the posterior mean \hat{b}_T , which, by Theorem 3.3, is the unique weak solution in $\dot{H}^p(\mathbb{T})$ of the differential equation (3.4).

Theorem 5.1. *Suppose that the true drift function $b_0 \in \dot{H}^p(\mathbb{T})$. Then for every $\delta > 0$,*

$$\|\hat{b}_T - b_0\|_{L^2} = O_{\mathbb{P}_{b_0}}(T^{-\frac{p-1/2}{2p}+\delta})$$

as $T \rightarrow \infty$.

Proof. By Theorem 3.3 we have (in the weak sense)

$$(\mathcal{C}_0^{-1} + L_T^\circ(\cdot; X))\hat{b}_T = \frac{1}{2}(L_T^\circ(\cdot; X))' + \chi_T^\circ(\cdot; X).$$

Note that it follows from (4.1) that ρ satisfies $\rho' = 2b_0\rho$ if b_0 is the drift function, hence, with $\mathbb{G}_T = \sqrt{T}(L_T^\circ(\cdot; X)/T - \rho)$,

$$\begin{aligned} (\mathcal{C}_0^{-1} + L_T^\circ(\cdot; X))b_0 &= \frac{1}{2}(L_T^\circ(\cdot; X))' + \chi_T^\circ(\cdot; X) + \mathcal{C}_0^{-1}b_0 \\ &\quad + \sqrt{T}\mathbb{G}_T b_0 - \frac{1}{2}\sqrt{T}\mathbb{G}_T' - \chi_T^\circ(\cdot; X). \end{aligned}$$

Subtracting the two equations shows that $e = \hat{b}_T - b_0$ satisfies (still in the weak sense)

$$(\mathcal{C}_0^{-1} + L_T^\circ(\cdot; X))e = -\mathcal{C}_0^{-1}b_0 - \sqrt{T}\mathbb{G}_T b_0 + \frac{1}{2}\sqrt{T}\mathbb{G}_T' + \chi_T^\circ(\cdot; X).$$

Since ρ is bounded away from zero (see (4.1)) and statement (i) of Theorem 4.1 says that, almost surely, L_T°/T converges uniformly on \mathbb{T} to ρ , it follows that there exists a constant $c > 0$ such that

$$\inf_{x \in \mathbb{T}} \frac{L_T^\circ(x; X)}{T} \geq c$$

on an event A_T with $\mathbb{P}_{b_0}(A_T) \rightarrow 1$. As a consequence, testing the weak differential equation for the error e with the test function e itself (“energy method”) yields, on the event A_T , the inequality

$$\begin{aligned} &\|\mathcal{C}_0^{-1/2}e\|_{L^2}^2 + cT\|e\|_{L^2}^2 \\ &\leq |\langle \mathcal{C}_0^{-1}b_0, e \rangle| + |\langle \chi_T^\circ(\cdot; X), e \rangle| + \sqrt{T}|\langle \mathbb{G}_T b_0, e \rangle| + \frac{1}{2}\sqrt{T}|\langle \mathbb{G}_T, e' \rangle| \\ &= |\langle \mathcal{C}_0^{-\frac{1}{2}}b_0, \mathcal{C}_0^{-\frac{1}{2}}e \rangle| + |\langle \chi_T^\circ(\cdot; X), e \rangle| + \sqrt{T}|\langle \mathbb{G}_T b_0, e \rangle| + \frac{1}{2}\sqrt{T}|\langle \mathbb{G}_T, e' \rangle|. \end{aligned}$$

We now use Young’s inequality ([29], Lemma 1.8) in the form

$$2\langle f, g \rangle \leq \kappa\|f\|_{L^2}^2 + \kappa^{-1}\|g\|_{L^2}^2$$

to estimate the first three terms on the right. Choosing appropriate κ ’s and subtracting the resulting terms involving $T\|e\|_{L^2}^2$ from both sides we get, still on A_T ,

$$\begin{aligned} &\|\mathcal{C}_0^{-1/2}e\|_{L^2}^2 + cT\|e\|_{L^2}^2 \\ &\lesssim \|\mathcal{C}_0^{-1}b_0\|_{L^2}^2 + \frac{\|\chi_T^\circ(\cdot; X)\|_{L^2}^2}{T} + \|b_0\|_\infty^2\|\mathbb{G}_T\|_{L^2}^2 + \frac{1}{2}\sqrt{T}|\langle \mathbb{G}_T, e' \rangle|. \end{aligned} \tag{5.1}$$

We note that the first three terms on the right are now stochastically bounded: the first one is constant, the second is bounded by a $|X_T - X_0|^2/T$, which is $O_{\mathbb{P}}(1)$ according to Corollary 4.2, and the third one is $O_{\mathbb{P}}(1)$ by Theorem 4.1.

For the last term on the right we have, since the norm $\|\mathcal{C}_0^{-s/(2p)} \cdot\|_{L^2}$ is equivalent to the H^s -norm and $\mathcal{C}_0^{1/(2p)} \frac{\partial}{\partial x}$ is bounded,

$$|\langle \mathbb{G}_T, e' \rangle| = \left| \left\langle \mathcal{C}_0^{-\frac{s}{2p}} \mathbb{G}_T, \mathcal{C}_0^{\frac{s}{2p}} e' \right\rangle \right| \lesssim \|\mathbb{G}_T\|_{H^s} \|\mathcal{C}_0^{\frac{s-1}{2p}} e\|_{L^2}.$$

We now use the interpolation inequality given as Theorem 13 on p.149 in [10],

$$\|A^\iota u\| \leq \|Au\|^\iota \|u\|^{1-\iota},$$

which is valid for $\iota \in (0, 1)$ and positive, coercive, self-adjoint densely defined operators A . We take $A = \mathcal{C}_0^{-\frac{1}{2}}$ and $\iota = (1-s)/p$ and combining with what we had above we get

$$\sqrt{T} |\langle \mathbb{G}_T, e' \rangle| \lesssim T^{-\frac{s-1}{2p}} \|\mathbb{G}_T\|_{H^s} \|\mathcal{C}_0^{-\frac{1}{2}} e\|_{L^2}^{\frac{1-s}{p}} \|\sqrt{T} e\|_{L^2}^{\frac{p+s-1}{p}}.$$

Using Young's inequality again we have the further bound

$$\sqrt{T} |\langle \mathbb{G}_T, e' \rangle| \lesssim \kappa T^{-\frac{s-1}{p}} \|\mathbb{G}_T\|_{H^s}^2 + \kappa^{-1} \|\mathcal{C}_0^{-\frac{1}{2}} e\|_{L^2}^2 + \kappa^{-1} T \|e\|_{L^2}^2.$$

If we combine this with (5.1), choose κ large enough and subtract $\kappa^{-1} \|\mathcal{C}_0^{-\frac{1}{2}} e\|_{L^2}^2 + \kappa^{-1} T \|e\|_{L^2}^2$ from both sides of the inequality we arrive at the bound

$$T \|e\|_{L^2}^2 \leq O_{\mathbb{P}}(1) + T^{-\frac{s-1}{p}} \|\mathbb{G}_T\|_{H^s}^2,$$

which holds on the event A_T . In view of Theorem 4.1 and since $s \in (0, 1/2)$ is arbitrary, this completes the proof. \square

Theorem 5.1 only concerns the posterior mean, but we can in fact show that the whole posterior distribution contracts around the true b_0 at the same rate. As usual, we say that the posterior contracts around b_0 at the rate ε_T (relative to the L^2 -norm) if for arbitrary positive numbers $M_T \rightarrow \infty$,

$$\mathbb{E}_{b_0} \Pi(b : \|b - b_0\|_{L^2} \geq M_T \varepsilon_T \mid X) \rightarrow 0$$

as $T \rightarrow \infty$. This essentially says that for large T , the posterior mass is concentrated in L^2 -balls around b_0 with a radius of the order ε_T .

Theorem 5.2. *Suppose that $b_0 \in \dot{H}^p(\mathbb{T})$. Then for every $\delta > 0$, the posterior contracts around b_0 at the rate $T^{-(p-1/2)(2p)+\delta}$ as $T \rightarrow \infty$.*

Proof. Set $\varepsilon_T = T^{-(p-1/2)(2p)+\delta}$ and consider arbitrary positive numbers $M_T \rightarrow \infty$. By the triangle inequality,

$$\begin{aligned} & \mathbb{E}_{b_0} \Pi(b : \|b - b_0\|_{L^2} \geq M_T \varepsilon_T \mid X) \\ & \leq \mathbb{E}_{b_0} \Pi\left(b : \|b - \hat{b}_T\|_{L^2} \geq \frac{M_T \varepsilon_T}{2} \mid X\right) + \mathbb{P}_{b_0}\left(\|\hat{b}_T - b_0\|_{L^2} \geq \frac{M_T \varepsilon_T}{2}\right). \end{aligned}$$

By Theorem 5.1 the second term on the right vanishes as $T \rightarrow \infty$, hence, since the posterior measure of a set is bounded by 1, it suffices to show that $\Pi(b : \|b - \hat{b}_T\|_{L^2} \geq M_T \varepsilon_T / 2 \mid X)$ converges to 0 in \mathbb{P}_{b_0} -probability. By Markov's inequality, this quantity is bounded by

$$\frac{4}{M_T^2 \varepsilon_T^2} \int \|\hat{b}_T - b\|_{L^2}^2 \Pi(db \mid X).$$

Since the integral is equal to the trace of the covariance operator of the centred posterior, it suffices to show that $\text{tr}(\mathcal{C}_T) = O_{\mathbb{P}_{b_0}}(\varepsilon_T^2)$. Since $M_T \rightarrow \infty$ this shows that that bound converges to zero in \mathbb{P}_{b_0} -probability.

As before, let λ_i and ϕ_i be the eigenvalues and eigenfunctions of the prior covariance operator \mathcal{C}_0 . For every $N \in \mathbb{N}$ we have

$$\text{tr}(\mathcal{C}_T) = \sum_{i \leq N} \langle \phi_i, \mathcal{C}_T \phi_i \rangle + \sum_{i > N} \langle \phi_i, \mathcal{C}_T \phi_i \rangle.$$

To bound the second sum on the right we note that in view of (3.3) we have $\mathcal{C}_T^{-1} \geq \mathcal{C}_0^{-1}$. Multiply this inequality by $\mathcal{C}_0^{1/2}$ from both sides to obtain $I + \mathcal{C}_0^{1/2} L_T^\circ \mathcal{C}_0^{1/2} \geq I$ and then, noting that $\mathcal{C}_0^{1/2} L_T^\circ \mathcal{C}_0^{1/2}$ is a bounded positive definite symmetric operator, multiply the inequality with $(I + \mathcal{C}_0^{1/2} L_T^\circ \mathcal{C}_0^{1/2})^{-1/2}$ on both sides to obtain $(I + \mathcal{C}_0^{1/2} L_T^\circ \mathcal{C}_0^{1/2})^{-1} \leq I$. Finally multiply both sides by $\mathcal{C}_0^{\frac{1}{2}}$ to arrive at $\mathcal{C}_T \leq \mathcal{C}_0$. Naturally, care has to be taken with the domain of definition of the unbounded operators involved, but first performing the calculations for the Fourier basis functions ϕ_k , one can pass to the limit, exploiting that each multiplication by $\mathcal{C}_0^{\frac{1}{2}}$ only adds compactness, see also Exercise 8, p.243 of [9] and the treatment in that chapter for more details.

Hence, since $\lambda_i \sim i^{-2p}$, the second sum is bounded by a constant times N^{1-2p} . By Cauchy-Schwarz the first sum is bounded by $\sum_{i \leq N} \|\mathcal{C}_T \phi_i\|_{L^2}$. To further bound this, we observe that

$$\begin{aligned} \inf_{x \in \mathbb{T}} L_T^\circ(x; X) \|\mathcal{C}_T \phi_i\|_{L^2}^2 &\leq \int_0^1 (\mathcal{C}_T \phi_i(x))^2 L_T^\circ(x; X) dx \\ &\leq \int_0^1 \mathcal{C}_T \phi_i(x) (\mathcal{C}_0^{-1} + L_T^\circ(\cdot; X)) \mathcal{C}_T \phi_i(x) dx \\ &= \int_0^1 \phi_i(x) \mathcal{C}_T \phi_i(x) dx \\ &\leq \|\phi_i\|_{L^2} \|\mathcal{C}_T \phi_i\|_{L^2} = \|\mathcal{C}_T \phi_i\|_{L^2}. \end{aligned}$$

Dividing by $\|\mathcal{C}_T \phi_i\|_{L^2}$ shows that $\|\mathcal{C}_T \phi_i\|_{L^2} \leq 1 / \inf_{x \in \mathbb{T}} L_T^\circ(x; X)$ and hence, by the first statement of Theorem 4.1, $\|\mathcal{C}_T \phi_i\|_{L^2} = O_{\mathbb{P}_{b_0}}(1/T)$.

Combining what we have we see that $\text{tr}(\mathcal{C}_T) \leq N O_{\mathbb{P}_{b_0}}(1/T) + N^{1-2p}$ for every $N \in \mathbb{N}$. The choice $N \sim T^{1/(2p)}$ balances the two terms and shows that $\text{tr}(\mathcal{C}_T) = O_{\mathbb{P}_{b_0}}(T^{(1-2p)/(2p)}) = O_{\mathbb{P}_{b_0}}(\varepsilon_T^2)$. \square

Remarks 5.3. *It is clear from the proof of Theorem 5.2 that the posterior spread $\int \|\hat{b}_T - b_0\|_{L^2}^2 \Pi(db | X)$ is always of the order $T^{(1-2p)/(2p)}$, regardless of the smoothness of the true drift function b_0 . Hence if the rate result of Theorem 5.1 for the posterior mean can be improved, for instance the condition that $b_0 \in \dot{H}^p(\mathbb{T})$ can be relaxed to the assumption $b_0 \in \dot{H}^{p-1/2}(\mathbb{T})$ (see the discussion in the concluding section), or the δ can be removed from the rate, then the result of Theorem 5.2 for the full posterior automatically improves as well.*

We also note that the proof of Theorem 5.1 delivers convergence rates in other norms. In particular it yields

$$\|\hat{b}_T - b_0\|_{H^p} = O_{\mathbb{P}_{b_0}}(T^{\frac{1}{4p} + \delta})$$

and hence, by interpolation ([29], Lemma 3.27) we have that the error in the mean converges to zero as

$$\|\hat{b}_T - b_0\|_{H^s} = O_{\mathbb{P}_{b_0}}(T^{\frac{1-2(p-s)}{4p} + \delta})$$

for $0 \leq s < p - 1/2$.

6 Proof of Theorem 4.1

6.1 Semi-martingale versus diffusion local time

Throughout this whole Section 6, the drift function $b \in \dot{C}(\mathbb{T})$ is fixed and, contrary to our use in previous sections, we denote the underlying law by \mathbb{P}_x when the diffusion is started in x and we sometimes shorten this to just \mathbb{P} when the diffusion is started in 0.

The weak solution X of the SDE (2.1) is a regular diffusion on \mathbb{R} with scale function s given by

$$s(x) = \int_{x_0}^x e^{-2 \int_{y_0}^y b(z) dz} dy.$$

We choose x_0 and y_0 such that $s(0) = 0$ and $s(1) = 1$. The speed measure m has Lebesgue density $1/s'$. Since b is 1-periodic and mean-zero the function s' is 1-periodic as well. It follows that m is 1-periodic and that s satisfies

$$s(x+k) = s(x) + k, \tag{6.1}$$

for all $x \in \mathbb{R}$ and $k \in \mathbb{Z}$.

The periodic local time L° was defined through the semi-martingale local time L of the diffusion X , for which we have the occupation times formula (2.3). The diffusion X also has continuous local time relative to its speed measure, the so-called diffusion local time of X . We denote this random field by $(\ell_t(x) : t \geq 0, x \in \mathbb{R})$. It holds that $t \mapsto \ell_t(x)$ is continuous and for every $t \geq 0$ and bounded, measurable function f ,

$$\int_0^t f(X_u) du = \int_{\mathbb{R}} f(x) \ell_t(x) m(dx) \tag{6.2}$$

(see for instance [20]). For this local time we define a periodic version ℓ° as well, by setting

$$\ell_t^\circ(x) = \sum_{k \in \mathbb{Z}} \ell_t(x+k).$$

The periodicity of m then implies that for every 1-periodic, bounded, measurable function f ,

$$\int_0^t f(X_u) du = \int_0^1 f(x) \ell_t^\circ(x) m(dx). \quad (6.3)$$

Comparing this with (2.5) we see that we have the relation $s'(x)L_T^\circ(x; X) = \ell_T^\circ(x)$ for every $T \geq 0$ and $x \in [0, 1]$. Now note that $1/s'$ is up to a constant equal to the invariant density ρ defined by (4.1). Since ρ is a probability density on $[0, 1]$ and $1/s'$ is the density of the speed measure m , we have $m[0, 1]\rho = 1/s'$ on $[0, 1]$. Therefore, statement (i) of Theorem 4.1 is equivalent to the statement that

$$\sup_{x \in \mathbb{T}} \left| \frac{1}{t} \ell_t^\circ(x) - \frac{1}{m[0, 1]} \right| \rightarrow 0 \quad (6.4)$$

a.s. as $t \rightarrow \infty$, and statement (ii) is equivalent to the asymptotic tightness of

$$x \mapsto \sqrt{t} \left(\frac{1}{t} \ell_t^\circ(x) - \frac{1}{m[0, 1]} \right) \quad (6.5)$$

in $H^\alpha(\mathbb{T})$ for every $\alpha \in [0, 1/2]$. We will prove these statements in the subsequent subsections.

6.2 A representation of the local time up to winding times

We define a sequence of \mathbb{P}_0 -a.s. finite stopping times τ_0, τ_1, \dots by setting $\tau_0 = 0$, τ_1 is the first time X exits $[-1, 1]$, τ_2 is the first time after τ_1 that X exits $[X_{\tau_1} - 1, X_{\tau_1} + 1]$, etc. (Note that if we define a process Z on the complex unit circle by $Z_t = \exp(2i\pi X_t)$, then τ_k is the time that the process Z completes its k th winding of the circle.)

The following theorem gives a representation for the periodic local time of X up till the n th winding time. The representation involves a stochastic integral relative to $s(X)$. The process $s(X)$ is a diffusion in natural scale, hence a time-changed Brownian motion, and hence a continuous local martingale.

Theorem 6.1. *For $x \in (0, 1)$,*

$$\frac{1}{n} \ell_{\tau_n}^\circ(x) - 1 = \frac{1}{n} \sum_{k=1}^n U_k(x),$$

where U_1, \dots, U_n are i.i.d. continuous random functions, distributed as

$$U(x) = \ell_{\tau_1}(x) + \ell_{\tau_1}(x-1) - 1 \quad (6.6)$$

$$= X_{\tau_1}(1 - 2s(x)) + 2 \int_0^{\tau_1} \phi_x(X_u) ds(X_u), \quad (6.7)$$

where $\phi_x = 1_{(x-1, \infty)} - 1_{(-\infty, x]}$.

Proof. For $k \in \mathbb{N}$ we write $X^k = (X_{\tau_{k-1}+t} - X_{\tau_{k-1}} : t \geq 0)$ and $\tau_1^k = \inf\{t : |X_t^k| = 1\}$. By Lemma 6.2 ahead, the processes $(X_t^k : t \in [0, \tau_1^k])$ are independent and have the same distribution as $(X_t : t \in [0, \tau_1])$. It follows that for $x \in (0, 1)$, with ℓ^Z denoting the diffusion local time of the diffusion Z ,

$$\frac{1}{n} \ell_{\tau_n}^{\circ}(x) - 1 = \frac{1}{n} \sum_{k=1}^n U_k(x), \quad (6.8)$$

where

$$U_k(x) = \ell_{\tau_1^k}^{X^k}(x) + \ell_{\tau_1^k}^{X^k}(x-1) - 1,$$

and the U_k are independent copies of the random function U defined by (6.6).

Now let $Y = s(X)$. Then Y is a regular diffusion in natural scale (i.e. the identity function is its scale function) and the speed measure m^Y of Y is related to the speed measure m of X by $m = m^Y \circ s$. It is easily seen that for ℓ^Y the local time of Y relative to its speed measure m^Y , we have $\ell_t^Y(x) = \ell_t^Y(s(x))$. For diffusions in natural scale, the diffusion local time coincides with the semimartingale local time (see [30], Section V.49). In particular, the Tanaka-Méyer formula holds:

$$\ell_t^Y(x) = |Y_t - x| - |x| - \int_0^t \text{sign}(Y_u - x) dY_u \quad (6.9)$$

under \mathbb{P}_0 . In view of (6.1) τ_1 is also the first time that Y exits $[-1, 1]$, so we have that $X_{\tau_1} = Y_{\tau_1}$. Using also the fact that the scale function s is strictly increasing, we obtain

$$\ell_{\tau_1}(x) = |X_{\tau_1} - s(x)| - |s(x)| - \int_0^{\tau_1} \text{sign}(X_u - x) ds(X_u). \quad (6.10)$$

Together with (6.1) this implies that (6.7) holds. \square

The proof of the theorem uses the following lemma, which implies that X “starts afresh” after every winding time τ_k . Let $(\mathcal{F}_t : t \geq 0)$ denote the natural filtration of the process X .

Lemma 6.2. *For every \mathbb{P}_0 -a.s. finite stopping time τ such that $X_\tau \in \mathbb{Z}$ a.s., it holds that the process $(X_{\tau+t} - X_\tau : t \geq 0)$ is independent of \mathcal{F}_τ and has the same law as X under \mathbb{P}_0 .*

Proof. Fix a measurable subset $C \subset C[0, \infty)$. By the strong Markov property we have

$$\mathbb{P}_0(X_{\tau+} - X_\tau \in C | \mathcal{F}_\tau) = f(X_\tau)$$

a.s., where $f(x) = \mathbb{P}_x(X - X_0 \in C)$. The periodicity of the drift function implies that for every $k \in \mathbb{Z}$, $f(k) = \mathbb{P}_k(X - k \in C) = \mathbb{P}_0(X \in C)$. Hence we have

$$\mathbb{P}_0(X_{\tau+} - X_\tau \in C | \mathcal{F}_\tau) = \mathbb{P}_0(X \in C),$$

a.s., which completes the proof. \square

Since we will be interested in the local time up till a deterministic time t , it is necessary to deal with the time interval between t and the previous or next winding time. The following lemma will be used for that. For $t \geq 0$, let the \mathbb{Z}_+ -valued random variable n_t be such that τ_{n_t} is the last winding time less or equal to t , so $\tau_{n_t} \leq t < \tau_{n_t+1}$.

Lemma 6.3. *For all $t \geq 0$ and Borel sets $B \subset C[0, 1]$,*

$$\mathbb{P}_0(\ell_{\tau_{n_t}+1}^\circ - \ell_t^\circ \in B) = \mathbb{E}_0 \mathbb{P}_{X_t - X_{n_t}}(\ell_{\tau_1}^\circ \in B).$$

Proof. We split up the event of interest according to the position of X at time τ_{n_t} . For $k \in \mathbb{Z}$ we have

$$\mathbb{P}_0(\ell_{\tau_{n_t}+1}^\circ - \ell_t^\circ \in B; X_{\tau_{n_t}} = k) = \mathbb{P}_0(\ell_{\sigma_{t,k}}^\circ - \ell_t^\circ \in B; X_{\tau_{n_t}} = k),$$

where $\sigma_{t,k} = \inf\{s > t : |X_s - k| \geq 1\}$. Let $(\mathcal{F}_s : s \geq 0)$ be the natural filtration of the process X . Since $X_{\tau_{n_t}}$ is \mathcal{F}_t -measurable, conditioning on \mathcal{F}_t gives

$$\mathbb{P}_0(\ell_{\tau_{n_t}+1}^\circ - \ell_t^\circ \in B; X_{\tau_{n_t}} = k) = \mathbb{E}_0 1_{\{X_{\tau_{n_t}}=k\}} \mathbb{P}_0(\ell_{\sigma_{t,k}}^\circ - \ell_t^\circ \in B | \mathcal{F}_t).$$

By the Markov property, the conditional probability equals $\mathbb{P}_{X_t}(\ell_{\sigma_{0,k}}^\circ \in B)$. By the periodicity of the drift function, this is equal to $\mathbb{P}_{X_t-k}(\ell_{\sigma_{0,0}}^\circ \in B)$. Since $\sigma_{0,0} = \tau_1$, we obtain

$$\mathbb{P}_0(\ell_{\tau_{n_t}+1}^\circ - \ell_t^\circ \in B; X_{\tau_{n_t}} = k) = \mathbb{E}_0 1_{\{X_{\tau_{n_t}}=k\}} \mathbb{P}_{X_t-k}(\ell_{\tau_1}^\circ \in B).$$

Summation over k completes the proof. \square

6.3 Proof of statement (i) of Theorem 4.1

In this subsection we prove that (6.4) holds a.s. for $T \rightarrow \infty$, which is equivalent to statement (i) of Theorem 4.1.

According to Theorem 6.1 we have

$$\frac{1}{n} \ell_{\tau_n}^\circ(x) - 1 = \frac{1}{n} \sum_{k=1}^n U_k(x),$$

where the U_k are independent copies of the continuous random function on $[0, 1]$ given by (6.6). Now $\mathbb{E}\|U\|_\infty \leq 1 + 2\mathbb{E} \sup_{|x| \leq 1} \ell_{\tau_1}(x)$. To bound the expectation, we again use the fact that $\ell_{\tau_1}(x) = \ell_{\tau_1}^Y(s(x))$, for $Y = s(X)$. Relation (6.1) implies that $\sup_{|x| \leq 1} \ell_{\tau_1}(x) = \sup_{|x| \leq 1} \ell_{\tau_1}^Y(x)$. Applying the BDG-type inequality for local times to the stopped continuous local martingale Y^{τ_1} (see [28], Theorem XI.(2.4)) we then see that for some constant $C > 0$,

$$\mathbb{E}\|U\|_\infty \leq 1 + 2\mathbb{E} \sup_{|x| \leq 1} \ell_{\tau_1}^Y(x) \leq 1 + C\mathbb{E} \sup_{t \leq \tau_1} |Y_t| < \infty.$$

Since by (6.1) it holds that $X_{\tau_1} = \pm 1$ with equal probability, it easily derived from (6.7) that $\mathbb{E}U(x) = 0$. By the Banach space version of Kolmogorov's law of large numbers (see [23], Corollary 7.10), it follows that

$$\sup_{x \in [0,1]} \left| \frac{1}{n} \ell_{\tau_n}^\circ(x) - 1 \right| \rightarrow 0 \quad (6.11)$$

a.s..

The random variables $\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_2, \dots$ are i.i.d., so by the law of large numbers, $\tau_n/n \rightarrow \mathbb{E}\tau_1$ a.s.. Applying relation (6.2) with $t = \tau_1$ and $f \equiv 1$ we see that

$$\mathbb{E}\tau_1 = \int_{-1}^1 \mathbb{E}\ell_{\tau_1}(x) m(dx).$$

Since $X_{\tau_1} = \pm 1$ with equal probability, (6.10) implies that $\mathbb{E}\ell_{\tau_1}(x) = 1 - |s(x)|$. Using (6.1) and the periodicity of m , it follows that

$$\begin{aligned} \mathbb{E}\tau_1 &= \int_{-1}^0 (1 + s(x)) m(dx) + \int_0^1 (1 - s(x)) m(dx) \\ &= \int_0^1 (1 + s(x-1)) m(dx) + \int_0^1 (1 - s(x)) m(dx) = m[0,1]. \end{aligned}$$

Combining (6.11) with the fact that $\tau_n/n \rightarrow m[0,1]$ a.s., we find that

$$\sup_{x \in [0,1]} \left| \frac{1}{\tau_n} \ell_{\tau_n}^\circ(x) - \frac{1}{m[0,1]} \right| \rightarrow 0 \quad (6.12)$$

a.s..

Now let n_t be defined as before Lemma 6.3, so that $\tau_{n_t} \leq t < \tau_{n_t+1}$. Then as $t \rightarrow \infty$ it holds that $n_t \rightarrow \infty$ and hence $\tau_{n_t}/n_t \rightarrow m[0,1]$ a.s. and $\tau_{n_t+1}/n_t \rightarrow m[0,1]$ a.s.. It follows that $n_t/t \rightarrow 1/m[0,1]$ a.s., and therefore also $\tau_{n_t}/t \rightarrow 1$ a.s.. We can write

$$\frac{1}{t} \ell_t^\circ(x) = \frac{\tau_{n_t}}{t} \frac{1}{\tau_{n_t}} \ell_{\tau_{n_t}}^\circ(x) + \frac{1}{t} (\ell_t^\circ(x) - \ell_{\tau_{n_t}}^\circ(x)).$$

Relation (6.12) shows that a.s., the first term on the right converges uniformly to $1/m[0,1]$. The second term is non-negative and bounded by

$$\frac{1}{t} (\ell_{\tau_{n_t}+1}^\circ(x) - \ell_{\tau_{n_t}}^\circ(x)) = \frac{\tau_{n_t+1}}{t} \frac{1}{\tau_{n_t+1}} \ell_{\tau_{n_t+1}}^\circ(x) - \frac{\tau_{n_t}}{t} \frac{1}{\tau_{n_t}} \ell_{\tau_{n_t}}^\circ(x),$$

which converges uniformly to 0 by the preceding. This completes the proof of (6.4) and hence of statement (i) of Theorem 4.1.

6.4 Proof of statement (ii) of Theorem 4.1

In this subsection we prove that the random maps (6.5) are asymptotically tight in the space in $H^\alpha(\mathbb{T})$ for every $\alpha \in [0, 1/2)$, which is equivalent to statement (ii)

of Theorem 4.1. It is most convenient and of course not restrictive to work with the complex Sobolev spaces. Let $e_k(x) = \exp(i2k\pi x)$, $k \in \mathbb{Z}$, be the standard complex exponential basis of $L^2[0, 1]$. For $\alpha \geq 0$, define the associated Sobolev space

$$H^\alpha[0, 1] = \left\{ f \in L^2[0, 1] : \|f\|_{H^\alpha}^2 = \sum |k|^{2\alpha} |\langle f, e_k \rangle|^2 < \infty \right\},$$

where $\langle f, g \rangle = \int_0^1 f(x) \bar{g}(x) dx$ is the usual inner product on $L^2[0, 1]$.

By the representation of the local time given by Theorem 6.1 and the central limit theorem for Hilbert space-valued random elements (e.g. [23], Corollary 10.9), we have that

$$\sqrt{n} \left(\frac{1}{n} \ell_{\tau_n}^\circ - 1 \right) \quad (6.13)$$

converges weakly in $H^\alpha[0, 1]$ if

1. $\mathbb{E} \|U\|_{H^\alpha}^2 < \infty$.
2. $\mathbb{E} U = 0$ (where the expectation is to be interpreted as a Pettis integral),

We will show that these conditions hold if (and only if) $\alpha < 1/2$. Slightly abusing notation, denote the two functions on the right of (6.7) by U_1 and U_2 . We will show that conditions 1–2 hold for U_1 and U_2 separately.

To show that the conditions hold for U_1 , recall that $X_{\tau_1} \pm 1$ with equal probability. Hence, $\mathbb{E} U_1 = 0$ and $\mathbb{E} \|U_1\|_{H^\alpha}^2 = \|1 - 2s\|_{H^\alpha}^2 < \infty$.

As for U_2 , using (6.7) and the stochastic Fubini theorem it is readily checked that

$$\langle U_2, e_k \rangle = 2 \int_0^{\tau_1} c_k(X_u) ds(X_u),$$

where

$$c_k(x) = \begin{cases} \frac{1 - e^{i2k\pi}}{i2k\pi} & \text{if } x + 1 \leq 0, \\ \frac{e^{i2k(x+1)\pi} - e^{i2k\pi}}{i2k\pi} & \text{if } x \leq 0 \leq x + 1 \leq 1, \\ \frac{e^{i2k\pi x} - 1}{i2k\pi} & \text{if } 0 \leq x \leq 1 \leq x + 1, \\ \frac{e^{i2k\pi} - 1}{i2k\pi} & \text{if } x \geq 1. \end{cases}$$

To show that condition 1. holds for U_2 , note that for $u \leq \tau_1$ it holds that $|X_u| \leq 1$. It is straightforward to see that for $|x| \leq 1$, we have $|c_k(x)| \leq C(1 + |k|)^{-1}$ for some $C > 0$. Therefore, by the Itô isometry,

$$\begin{aligned} \mathbb{E} \sum |k|^{2\alpha} \left| \int_0^{\tau_1} c_k(X_u) ds(X_u) \right|^2 &= \sum |k|^{2\alpha} \mathbb{E} \int_0^{\tau_1} |c_k(X_u)|^2 d\langle s(X) \rangle_u \\ &\leq C^2 \mathbb{E} \langle s(X) \rangle_{\tau_1} \sum \frac{|k|^{2\alpha}}{(1 + |k|)^2}. \end{aligned}$$

The sum on the right is finite if $\alpha \in [0, 1/2)$. For the diffusion $Y = s(X)$ the diffusion local time coincides with the semi-martingale local time, hence

$$\mathbb{E} \langle s(X) \rangle_{\tau_1} = \mathbb{E} \int \ell_{\tau_1}^Y(x) dx = \int_{-1}^1 \mathbb{E} \ell_{\tau_1}^Y(x) dx.$$

The Tanaka-Méyer formula and optional stopping imply that for $|x| \leq 1$,

$$\mathbb{E} \ell_{\tau_1}^Y(x) = \mathbb{E} |Y_{\tau_1} - x| - |x| = 1 - |x|.$$

Hence, U_2 satisfies condition 1. Finally, note that to show that $\mathbb{E} U_2 = 0$, it suffices to show that $\mathbb{E} U_2(x) = 0$ for every fixed $x \in (0, 1)$. But this follows readily from (6.6) again, by optional stopping. So indeed the random maps (6.13) converge in $H^\alpha[0, 1]$ for every $\alpha \in [0, 1/2)$.

To complete the proof we consider the decomposition

$$\begin{aligned} & \sqrt{t} \left(\frac{\ell_t^\circ}{t} - \frac{1}{m[0, 1]} \right) \\ &= \sqrt{n_t + 1} \left(\frac{\ell_{\tau_{n_t+1}}^\circ}{n_t + 1} - 1 \right) \sqrt{\frac{n_t + 1}{t}} + \sqrt{t} \left(\frac{n_t + 1}{t} - \frac{1}{m[0, 1]} \right) - \frac{\ell_{\tau_{n_t+1}}^\circ - \ell_t^\circ}{\sqrt{t}}. \end{aligned}$$

Since $n_t/t \rightarrow 1/m[0, 1]$ a.s., the tightness of the maps (6.13) implies that the first term is asymptotically tight. By the central limit theorem, $\sqrt{n}(\tau_n/n - m[0, 1])$ converges in distribution. Together with the inequality $\tau_{n_t} \leq t < \tau_{n_t+1}$ and the delta method this implies that the second term is asymptotically tight as well. For the last term, note that by Lemma 6.3 we have, for $M > 0$,

$$\mathbb{P}_0 \left(\left\| \frac{\ell_{\tau_{n_t+1}}^\circ - \ell_t^\circ}{\sqrt{t}} \right\|_{H^\alpha} > M \right) \leq \sup_{|a| \leq 1} \mathbb{P}_a(\|\ell_{\tau_1}\|_{H^\alpha} > M\sqrt{t}) \leq \frac{1}{M^2 t} \sup_{|a| \leq 1} \mathbb{E}_a \|\ell_{\tau_1}\|_{H^\alpha}^2.$$

Similar considerations as used to show that condition 1. above holds for U_2 show that the supremum over a on the right-hand side is bounded. We conclude that the last term in the decomposition is $o_P(1)$. This completes the proof.

7 Concluding remarks

We have obtained the posterior contraction rate $T^{-(p-1/2)/(2p)}$ for our non-parametric Bayes procedure. We remarked that the regularity of the prior is essentially $p - 1/2$ (Lemma 2.2) and assumed that the true drift b_0 has Sobolev regularity of order p . Although lower bounds for the rate of convergence in the exact model under study do not appear to be known, comparison with similar models suggests that the optimal rate for estimating a drift function b_0 that is β -regular (in Sobolev sense) may be $T^{-\beta/(1+2\beta)}$ in our setting (in a minimax sense over Sobolev balls for instance, cf. e.g. [22], [35] for similar results). The general message from the Gaussian process prior literature is that this optimal rate is typically attained if the “regularity” α of the prior matches the regularity

β of the function that is being estimated (see [40]). Since the regularity of the prior we employ in this paper is essentially $\alpha = p - 1/2$, this suggests that in principle, it should be possible to relax our assumption that b_0 is p -regular to the assumption that b_0 is $(p - 1/2)$ -regular, while still maintaining the same rate $T^{-(p-1/2)/(2p)}$. It is however not clear whether this can be achieved by adapting the proof we give in this paper. The method of proof is adapted from [1] where it is used to study linear inverse problems in the small noise limit. In that context the proof gives sharp rates in some parameter regimes, but not in others.

There are a number of future directions that this work could be taken in. First of all, alternative technical approaches could be explored to derive sharp convergence rates. One approach could be to use the representation of the posterior mean as a minimizer of some stochastic objective functional (cf. Subsection 3.3) and use empirical process-type techniques to study its asymptotic properties. This however requires technical tools (e.g. uniform limit theorems, maximal inequalities) that are presently not available in this setting of periodic diffusions. Alternatively, sharp rates may result from a general rate of convergence theory for posteriors in the spirit of [36], if that could be developed for this class of models. Secondly, motivated by practical considerations, it will also be interesting to determine whether useful adaptive procedures can be constructed by choosing the hyper-parameters p, η and κ in a data-driven way, for instance by hierarchical Bayes or empirical Bayes procedures. There is recent computational work in this direction, cf. [32], but no theoretical results are presently available. A third future direction concerns extension of the ideas in this paper to diffusions in more than one dimension. The local time is, then, a much more singular object and developing an analysis of posterior consistency will present new challenges.

Acknowledgements

The research of AMS is funded by the EPSRC (UK) and by the ERC. The research of JHvZ is supported by the Netherlands Organization for Scientific Research NWO. The authors are grateful to Sergios Agapiou for helpful comments.

References

- [1] AGAPIOU, S., AND LARSSON, S., AND STUART, A.M. Posterior Consistency of the Bayesian Approach to Linear Ill-Posed Inverse Problems. Preprint: arXiv:1203.5753v2 [math.ST]
- [2] BANDI, F.M., AND PHILLIPS, P.C.B. Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71, 1 (2003), 241–283.
- [3] BANON, G. Nonparametric identification for diffusion processes. *SIAM J. Control Optim.* 16, 3 (1978), 380–395.

- [4] BHATTACHARYA, R. A Central Limit Theorem for Diffusions with Periodic Coefficients. *Ann. Probab.* 13, 2 (1985), 385–396.
- [5] BOGACHEV, V.I. *Gaussian Measures*. AMS, 1998.
- [6] BOLTHAUSEN, E. On the global asymptotic behavior of Brownian local time on the circle. *Trans. Amer. Math. Soc.* 253 (1979), 317–328.
- [7] CASTILLO, I. Lower bounds for posterior rates with Gaussian process priors. *Electron. J. Stat.* 2 (2008), 1281–1299.
- [8] COMTE, F., GENON-CATALOT, V., AND ROZENHOLC, Y. Penalized non-parametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* 13, 2 (2007), 514–543.
- [9] CONWAY, J. B. A Course in Functional Analysis. *Springer* (1990).
- [10] DAUTRAY, R., AND LIONS, J.L. *Mathematical Analysis and Numerical Methods for Science and Technology*, vol. 3. Springer, 1990.
- [11] DIACONIS, P., AND FREEDMAN, D. On the consistency of Bayes estimates. *Ann. Statist.* 14, 1 (1986), 1–67.
- [12] DURRETT, R. *Stochastic Calculus – A practical Introduction*. CRC Press, London, 1996.
- [13] EVANS, L.C. *Partial Differential Equations*. American Mathematical Society, 1998.
- [14] GHOSAL, S., GHOSH, J.K., AND VAN DER VAART, A.W. Convergence rates of posterior distributions. *Ann. Statist.* 28, 2 (2000), 500–531.
- [15] GHOSAL, S. AND VAN DER VAART, A.W. Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.* 35 1 (2007), 192223.
- [16] GHOSH, J.K. AND RAMAMOORTHY, R.V. *Bayesian nonparametrics*. Springer-Verlag, New York, 2003.
- [17] HAIRER, M., STUART, A.M., AND VOSS, J. Analysis of SPDEs arising in path sampling, part II: The nonlinear case. *Annals of Applied Probability* 17 (2007), 1657–1706.
- [18] HINDRIKS, R. *Empirical dynamics of neuronal rhythms*. PhD thesis, VU University Amsterdam, 2011.
- [19] IMKELLER, P., AND VON STORCH, J.-S., Eds. *Stochastic climate models* (Basel, 2001), vol. 49 of *Progress in Probability*, Birkhäuser Verlag.
- [20] ITÔ, K., AND MCKEAN, JR., H.P. *Diffusion processes and their sample paths*. Springer-Verlag, Berlin, 1974.

- [21] KARATZAS, I., AND SHREVE, S.E. *Methods of Mathematical Finance*. Springer, 1999.
- [22] KUTOYANTS, YU.A. *Statistical Inference for Ergodic Diffusion Processes*. Springer, 2004.
- [23] LEDOUX, M., AND TALAGRAND, M. *Probability in Banach Spaces*. Springer, 1991.
- [24] LIPTSER, R.S., AND SHIRYAEV, A.N. *Statistics of random processes. I*, expanded ed., Springer-Verlag, Berlin, 2001.
- [25] PANZAR, L. AND VAN ZANTEN, J.H. Nonparametric Bayesian inference for ergodic diffusions. *J. Statist. Plann. Inference* 139 12 (2009), 4193–4199.
- [26] PAPASPILIOPOULOS, O., POKERN, Y., ROBERTS, G.O., AND STUART, A.M. Nonparametric estimation of diffusions: a differential equations approach. *Biometrika*, doi:10.1093/biomet/ass034 (2012).
- [27] POKERN, Y. *Fitting Stochastic Differential Equations to Molecular Dynamics Data*. PhD thesis, University of Warwick, 2007.
- [28] REVUZ, D., AND YOR, M. *Continuous martingales and Brownian motion*, third ed., Springer-Verlag, Berlin, 1999.
- [29] ROBINSON, J. C. *Infinite-dimensional Dynamical Systems*, CUP, 2001.
- [30] ROGERS, L.C.G., AND WILLIAMS, D. *Diffusions, Markov processes, and martingales. Vol. 2*. John Wiley & Sons Inc., New York, 1987.
- [31] SAFONOVA, O. A. Asymptotic behavior of integral functionals of diffusion processes with periodic coefficients. *Ukrainian Math. J.* 44 2 (1992), 215–221.
- [32] SCHAUER, M., VAN DER MEULEN, F.H. AND VAN ZANTEN, J.H. Reversible jump MCMC for nonparametric drift estimation for diffusion processes. Preprint: arXiv:1206.4910v1 [stat.CO]
- [33] SCHLICK, T. *Molecular Modeling and Simulation, an Interdisciplinary Guide*. Springer, New York, 2002.
- [34] TARANTOLA, A. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2004.
- [35] TSYBAKOV, A.B. *Introduction to nonparametric estimation*. Springer, New York, 2009.
- [36] VAN DER MEULEN, F.H., VAN DER VAART, A.W. AND VAN ZANTEN, J.H. Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli* 12 5 (2006), 863–888.

- [37] VAN DER VAART, A.W., AND VAN ZANTEN, J.H. Donsker theorems for diffusions: necessary and sufficient conditions. *Ann. Probab.* *33*, 4 (2005), 1422–1451.
- [38] VAN ZANTEN, J.H. Rates of convergence and asymptotic normality of kernel estimators for ergodic diffusion processes. *J. Nonparametr. Statist.* *13*, 6 (2001), 833–850.
- [39] VAN ZANTEN, J.H. On uniform laws of large numbers for ergodic diffusions and consistency of estimators. *Stat. Inference Stoch. Process.* *6*, 2 (2003), 199–213.
- [40] VAN DER VAART, A.W., AND VAN ZANTEN, J.H. Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* *36* (2008), 1435–1463.
- [41] VAN DER VAART, A.W., AND VAN ZANTEN, J.H. Reproducing kernel Hilbert spaces of Gaussian priors. *IMS Collections 3* (2008), 200–222.
- [42] WAHBA, G. Spline models for observational data., Society for Industrial Mathematics, Philadelphia, 1990.